Comparison of Machine Learning Techniques for Predicting Online Shopping Dependency Level of Bangladeshi People

Md. Safiul Alam
Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

Abstract— **Bangladesh has one of the largest e-commerce markets. With the speed at which e-commerce is growing, it is becoming more and more important to understand and predict consumer behavior in this context. This research looks at four prominent machine learning techniques namely, logistic regression, k-nearest neighbor, support vector classifier, and multilayer perceptron, to see how well these classifiers predict the reliance on online purchasing. Several noteworthy performance measures are used to compare the performances of each classifier. Through an analysis of four algorithms and datasets pertinent to Bangladeshi internet shoppers, this study aims to shed light on the most effective machine learning approaches for evaluating the level of dependency on online shopping platforms. The study's findings broaden understanding of consumer behavior in Bangladeshi e-commerce and offer practical guidance to businesses engaged in this quickly evolving industry.**

Keywords— Bangladesh; comparison; dependency level; ecommerce; machine learning; online shopping

I. INTRODUCTION

In today's networked world, the growth of the internet has resulted in a fundamental shift in how we go about living our daily lives [1]. One of the clearest manifestations of this digital revolution is the exponential growth of e-commerce, which has changed the world economy and, quite impressively, permeated Bangladeshi culture [2]. Due to Bangladesh's quickly expanding middle class, rapid urbanization, and internet penetration, online shopping has grown to be a significant and dynamic component of modern consumer behavior [3]. One intriguing phenomenon that has caught the attention of academics, and businesses alike in this setting is the concept of online shopping dependence [4]. Online shopping has definitely enhanced convenience and access to a wide range of products and services, but it has also raised worries about the prospect of consumers being unduly dependent on online businesses for all of their purchasing requirements [5]. This reliance may have negative implications in some circumstances, such as increased financial stress, lower productivity, and even negative psychological effects [6].

For both the long-term growth of e-commerce in Bangladesh and the well-being of the general populace, it is crucial to comprehend the elements that affect dependency on online buying [7]. This context serves as the backdrop for the amazing comparison this work carries out in this study, a predictive model and comparison of four machine learning techniques that try to shed light on the numerous aspects of Bangladeshi consumers' dependency on online shopping and predict online shopping dependency level.

Machine learning is the study of utilizing computers to simulate human learning processes as well as the study of computer self-improvement methods for learning new skills and knowledge, recognizing existing knowledge, and consistently improving performance and success [8]. Algorithms for machine learning often ingest and analyze data in order to discover patterns relating to people, business processes, transactions, events, and so on [9]. The high failure rate of AI (Artificial Intelligence) processes and the focus on recommender systems show that machine learning algorithms have a lot of potential for application in e-commerce and it should be noted that in a field where businesses battle ferociously to stay competitive, the application of machine learning is still a relatively untapped area of study [10]. This work has used the power of machine learning algorithms to make a predictive model on a most important and unresearched topic to make Bangladeshi e-commerce services more effective which can increase the economic growth and financial inclusion in Bangladesh [11].

One of the main goals of this study has been accomplished, and that was to thoroughly assess and contrast four machine learning algorithms in terms of their capacity to anticipate and understand certain levels of reliance. By doing this, this research wants to further the academic discussion regarding online shopping practices and provide useful guidance to Bangladeshi businesses, officials, and consumer rights organizations.

Also, this research covers a wide range of elements that affect dependence on online shopping. In order to estimate and anticipate the dependence levels of Bangladeshi online buyers, this work has taken a data-driven strategy, utilizing large datasets and four machine learning methods.

Moreover, the overarching objective of this study is to contribute to the body of knowledge that directs Bangladesh's ecommerce-related decisions, legislation, and practices, ensuring that the benefits of the digital era are utilized while avoiding potential pitfalls [12].

In this research, four machine learning techniques have been applied for experimentation, and several performance evaluation metrics have been used to evaluate this unique piece of work. Four popular machine learning classifiers, including Logistic Regression, K-Nearest Neighbors (KNN), Support

Vector Classifier (SVC), and Multilayer Perceptron (MLP) have been experimented with on a survey dataset. Several performance evaluation metrics have been calculated to determine the best classifier in the working context, and a result comparison is presented here. From the analysis of the obtained result, it is confirmed that the K-Nearest Neighbors (KNN) classifier achieves the best result in terms of metrics.

These are the order of this paper: Section II gives an exhaustive overview of relevant studies. Section III provides the study methodology. Besides, result and discussion are provided in Section IV and a conclusion is given in Section V. Finally, Section VI provides future work

II. Literature Review

The many aspects of online shopping include locating online retailers and products, researching product information, selecting a payment method, connecting with other buyers and sellers, and making purchases of goods or services [13]. The increase in online shopping is part of a larger trend of rapidly growing Internet usage [14].

Shaffer et al. [15] offered an illustration of how coupons may be utilized as a strategic tool to attract clients and keep them loyal. Building on this idea, Dennis L. Dufy [16] found that e-commerce generates fresh opportunities with separate rules, particularly in terms of the growth of affiliate communities in both large and small networks of websites, resulting in a link to the market that benefits both sides.

Rust et al. [17] investigated the taxonomy of these new links across services in an information economy whereas Libai et al. [18] and Homburg et al. [19] focused on how the economics of the various affiliation mechanisms function, how they generate revenues, and how they bring short- and long-term advantages for businesses. Chen et al. [20] focused on the short- and long-term aspects of discounts, conversion rates, and short- and long-term profitability, whereas Jain et al. [21] investigated the business models of search engines as revenue generators for merchants and how these revenues are distributed.

Internet behemoths like Google, Amazon, and others started filing patents for techniques in the early 2010s that were based on the mathematical underpinnings of consumer networks and cashback websites [22]. They want to profit from their success or create the ideal business strategies for the merchants on such websites [23]. These methods served as the basis for fundamental research by Altinkemer et al. [24], Fu et al. [25], Ho et al. [26], and Roig-Tierno et al [27]. These studies highlighted the aforementioned traits and enabled empirical analysis of the usefulness and financial success of affiliate networks. These studies show that companies may increase their profits by building their social networks to increase traffic and referrals by offering incentives other than cashback.

Through empirical research, Vana et al. [28] focused on the financial viability of cashback websites and found that cashback payments increase the likelihood of purchases. Ballestar et al. [29], Ballestar et al. [30], and Ballestar et al. [31] made the case that it is essential to look at each enrolled customer's social media activity in order to understand their buying habits, business activity, loyalty, and most importantly long-term profitability. This is due to the fact that not every enrolled customer is active. Monaghan et al. [32] clarifies user interactions as well as the complex design of corporate and market institutions. While this study enables quantification of these connections, it is unable to predict future network behavior.

In order to support these assertions, Lamberton et al. [33] focused on the description and explanation of network dynamics, with the prediction of its changes acting as a critical test. Eggers et al. [34] stressed the importance of these dynamics, especially for entrepreneurial firms, which account for the bulk of the businesses in this sector. The demand for these organizations to contact a substantial number of customers in order to capitalize on the customer relationships that Kumar et al. [35] found is their major obstacle. This calls for the development and application of several evaluation processes.

The aim of this study is to evaluate and compare how well different machine learning approaches predict how reliant Bangladeshi consumers are on online shopping. Due to the intricacy of the research, the enormous amount of data, and the requirement to develop rigorous methods to evaluate the data, big data techniques like machine learning, artificial intelligence, and big data are beneficial to experimentally investigate this issue. The application of these technologies, in Harmeling et al. [36] opinion, enables the examination of unique user and business roles, interactions, and outcomes. This research objective has been condensed into the following hypothesis:

H1 A stunning comparison of machine learning methods to predict how dependent Bangladeshi customers are on online shopping; by identifying the most accurate model, businesses can better tailor their marketing strategies and services to the needs of Bangladeshi online purchasers.

III. Methodology

This section has been divided into four subsections which are Data Description, Classifier Description, Confusion Matrix and Classifier Evaluation Metrics, and Implementation Procedure. All of the details of the four subsections have been described below.

A. Data Description

To accomplish this unique work 1045 individual records have been used which had been collected by physical survey. In this research, a total of fourteen types of attributes of people have been used to carry out this work. Among the fourteen attributes, thirteen attributes (online shopping frequency, products bought online, preferred shopping platforms, frequent payment methods, influence convenience, influence price, influence variety, influence social media, influence recommendations, influence trust, influence access, cultural influence, primary shopping device) have been used as the explanatory variable and one attribute (online shopping dependency level) has been used as the response variable. Here 75% of the data from the total dataset has been used for training the classifiers and the rest 25% of the whole dataset has been used for testing the correctness of the classifiers.

B. Classifier Description

A classifier in machine learning is a tool for predicting the target characteristic from feature data points [37]. The theory that follows is applicable since four classifiers have been used to examine the dataset.

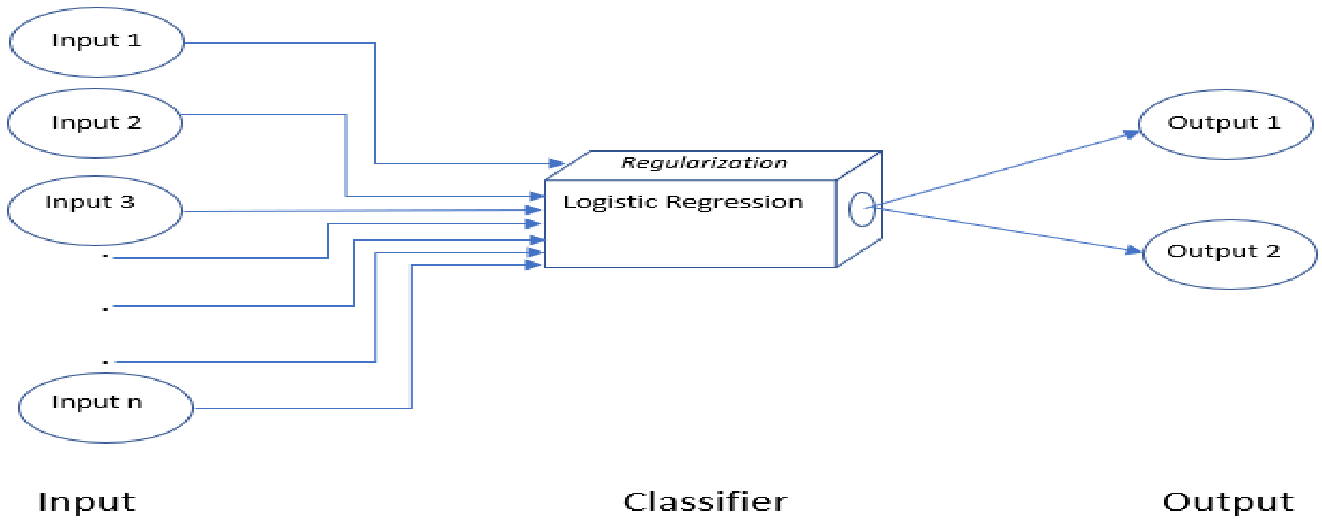

Fig. 1. Logistic Regression with Two Possible Outcomes.

Fig. 1 shows the working system of the Logistic Regression classifier which is a notorious supervised learning technique that has been used to predict the categorical dependent variable using a set of independent variables or features. This is a puissant and outstanding algorithm as it can give probabilities and classify new data by using both continuous and discrete data [38]. A logistic function is a mathematical formula that converts projected values into probabilities. Fig. 1. has also shown that this Logistic Regression can be classified into binomial types because there have been only two possible outcomes of the output variables.

The K-Nearest Neighbors (KNN) classifier is a straightforward and readily apparent supervised machine learning method for classification issues [39]. According to how similar they are to surrounding data points, it classifies the data points in a pattern space [40]. Selecting the value of "K", the number of nearest neighbors, and the distance metric used to determine similarity are crucial decisions have been shown in Fig. 2. When classifying a new data point, the method first determines the 'K' nearest neighbors, and then chooses the class name by majority vote [41]. The success of KNN depends on careful hyperparameter optimization and it is non-parametric and suitable for a wide variety of datasets [42].

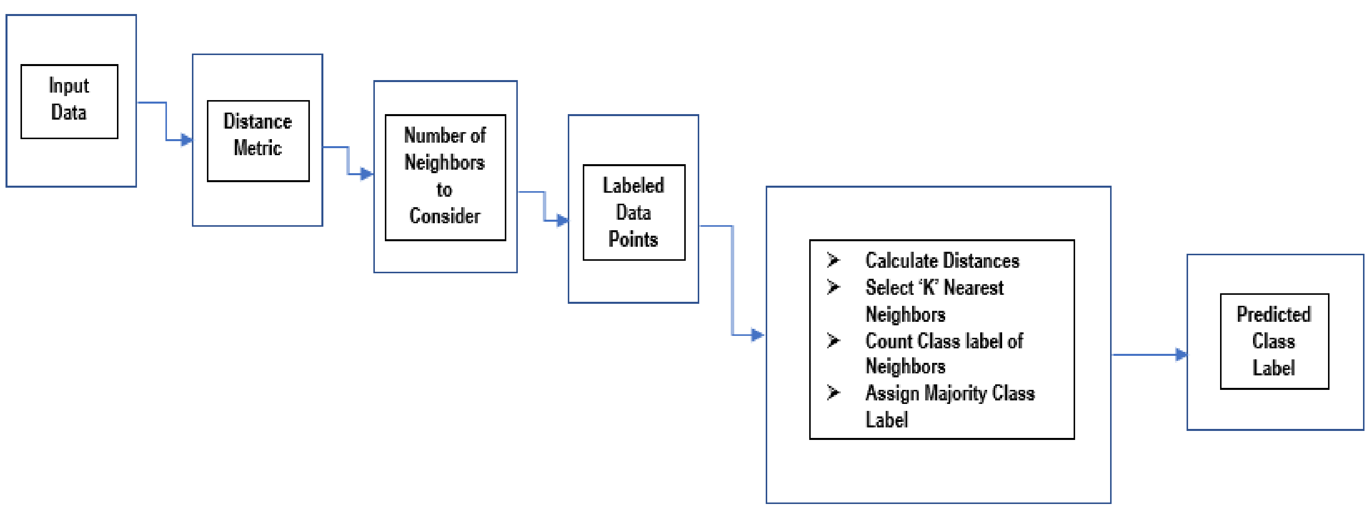

Fig. 2. Working Diagram of KNN.

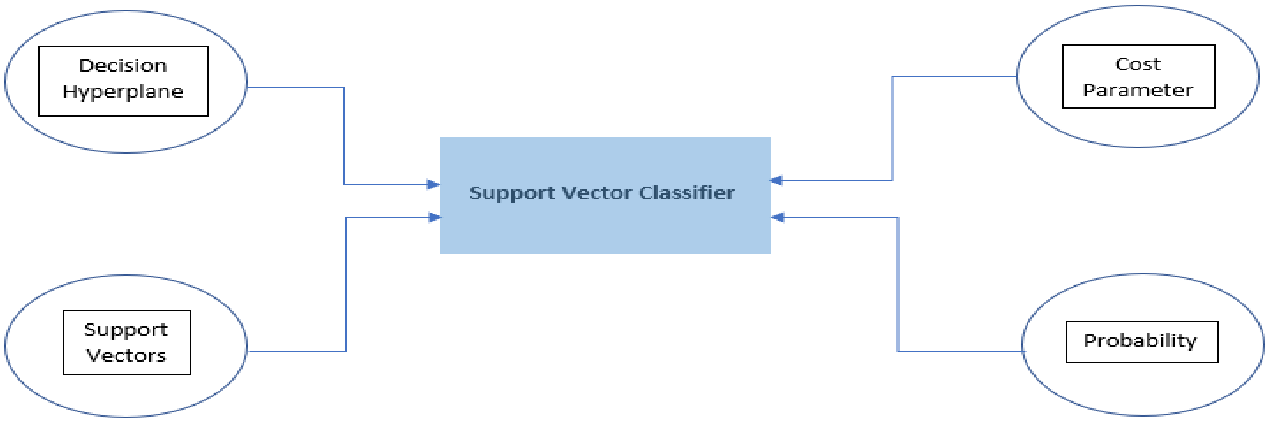

Fig. 3. Operational System of SVC.

The Support Vector Classifier (SVC) is a dependable supervised machine learning method for classification tasks and it also locates the optimal hyperplane in a high-dimensional feature space to classify data points [43]. The goal of SVC is to maximize the margin, which is a measurement of the distance between the hyperplane and the closest data points from each class even the margin optimization plays a key role in the model's good generalization to new, untested data [44]. SVC is adaptable and can handle both linear and non-linear classification problems by utilizing a range of kernel functions, including sigmoid, polynomial, linear, and radial basis function (RBF), and these kernels allow SVC to transform the input data into a higher-dimensional space, which facilitates the identification of complex, non-linear decision boundaries, because of its flexibility, the approach is particularly useful for tasks like text classification [45]. Its propensity for good generalization is indicated by its resistance to overfitting and outstanding performance in high-dimensional data sets [46]. To ensure optimal performance, however, proper hyperparameter adjustment is necessary which is connected to the operational system of SVC shown in Fig. 3. This process helps the SVC work well over a range of datasets and also aids in fine-tuning the model to prevent issues such as overfitting or underfitting [47].

Three different types of layers make up the multilayer perceptron, a feedforward neural network [48]. It also contains one or more hidden layers in addition to the input and output layers which has been shown in Fig. 4. In terms of practical applicability, a multilayer perceptron is seen to be superior to a single-layer perceptron [49]. A single-layer perceptron can only learn linear functions, but a multilayer perceptron can learn both linear and nonlinear functions [50]. Backpropagation is used to train the multilayer perceptron [51].

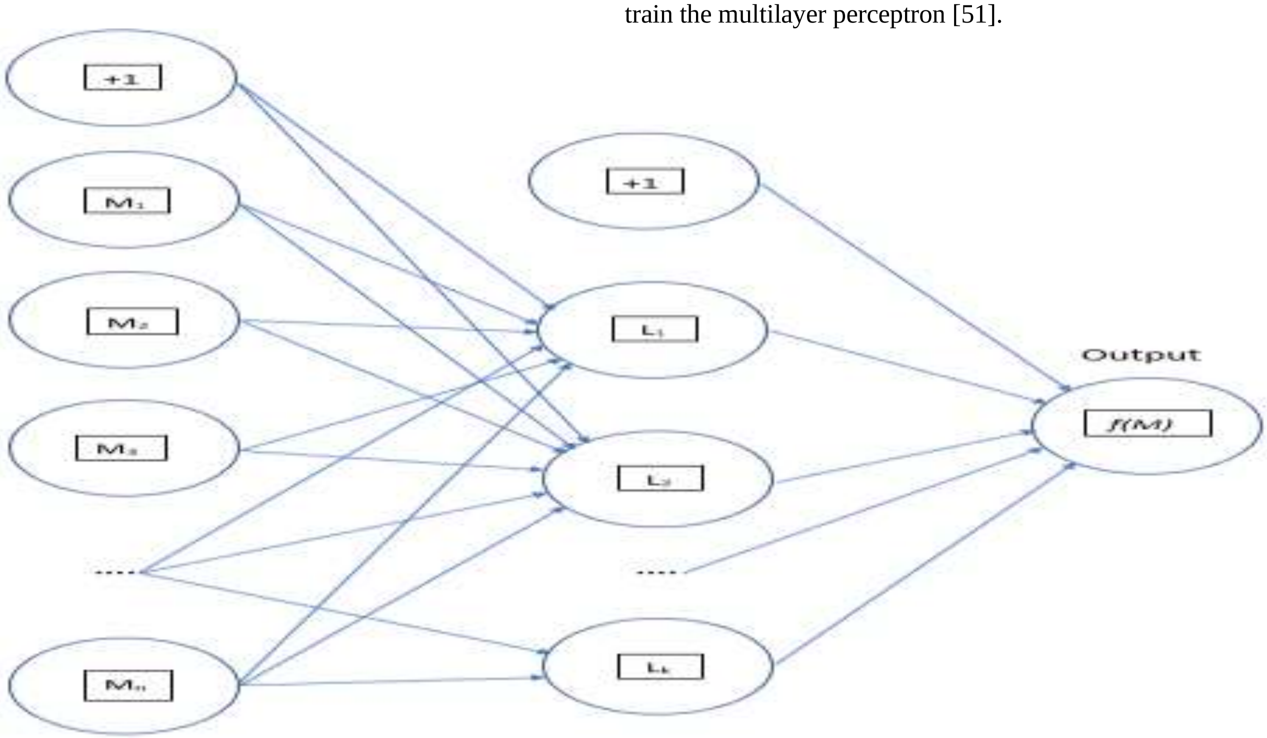

Fig. 4. One Hidden Layer in a Multilayer Perceptron.

C. Confusion Matrix and Classifier Evaluation Metrics

The confusion matrix is an indicator that indicates how many occasions a classification system accurately or mistakenly predicted [52]. This matrix may be used to evaluate a classifier's performance [53]. This particular problem involves two classes. The true positives, true negatives, false positives, and false negatives are all present in the case of the two-class issue. In this work, the true positive situation occurs when the model predicts a person's high dependency on online shopping and the actual output indicates a person's high dependency on online shopping. On the other hand, the true negative scenario occurs when the model predicts a person's low dependency on online shopping and the actual output indicates a person's low dependency on online shopping. Moreover, the false positive case occurs when the model predicts a person's high dependency on online shopping and the actual output indicates a person's low dependency on online shopping. On the contrary, the false positive issue occurs when the model predicts a person's low dependency on online shopping and the actual output indicates a person's high dependency on online shopping. A class classifier's capacity for prediction is evaluated using accuracy, recall, specificity, f1-score, and precision [54]. The confusion matrix was used to compute the performance evaluation metrics (Accuracy, Precision, Recall, F1-Score, Specificity False Positive Rate, and False Negative Rate) for the classifier [55].

D. Implementation Procedure

The implementation procedure of this work has been clearly described in Fig. 5. At first, the data has been collected through a physical survey then all the data has been stored for preprocessing. The label encoding technique is used for data preprocessing. Categorical columns may be converted to numerical ones via the Label Encoding method so that machine learning models that can only handle numerical data may fit them [56]. The dataset was methodically assembled using all of the respondents' replies to the 14 questions, and it was only after checking for null values that it was realized there were no missing values which has been shown in Fig. 6. Hereafter the data has been prepared. After that four most powerful machine learning algorithms namely Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Multilayer Perceptron (MLP) have been implemented. Consequently, those four algorithms have been trained with the help of 75% data and have been tested with the support of 25% data. Finally, the decision has been made with the help of a confusion matrix, classifier evaluation metrics, and comparative analysis.

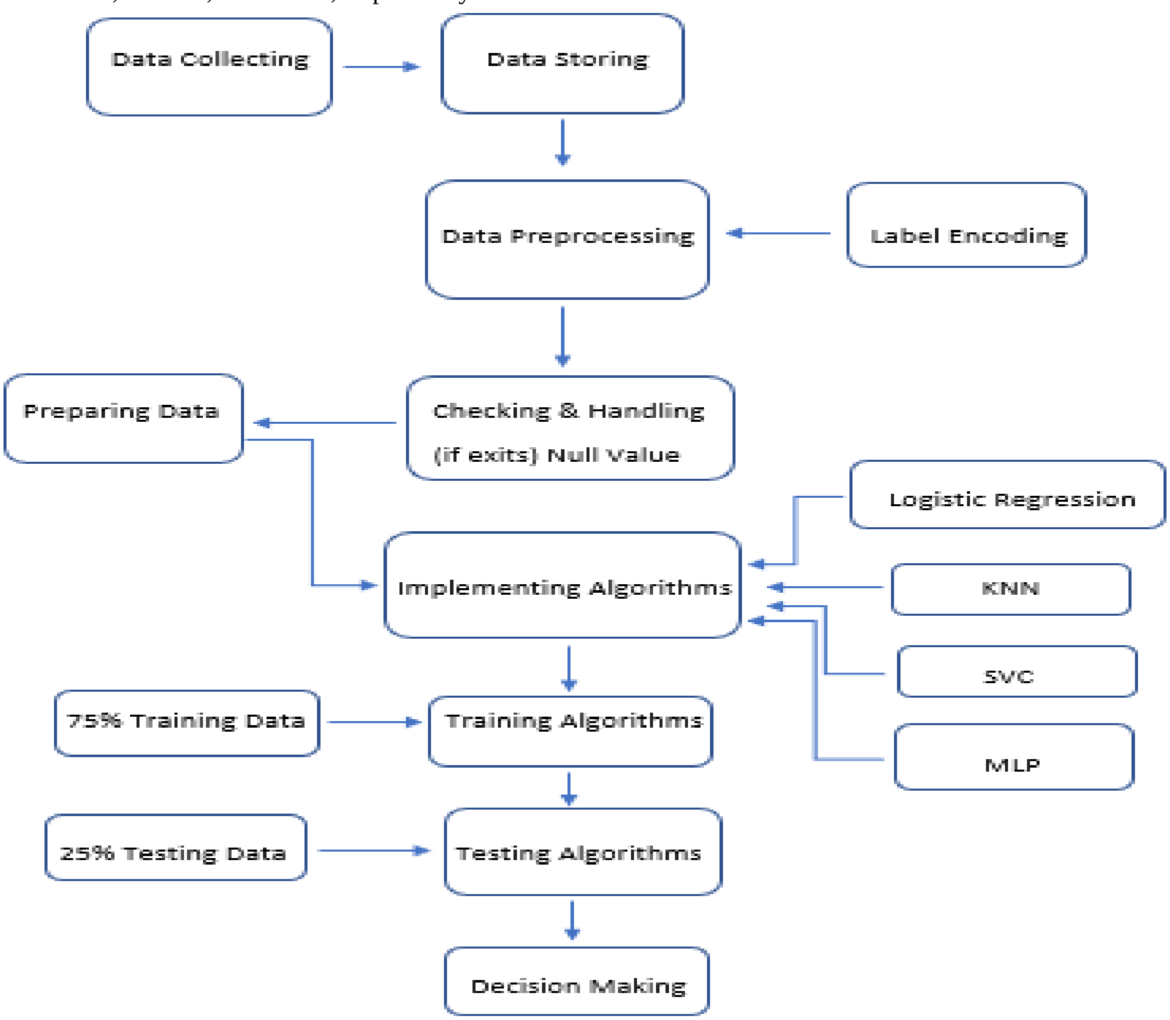

Fig. 5. Method for Implementation.

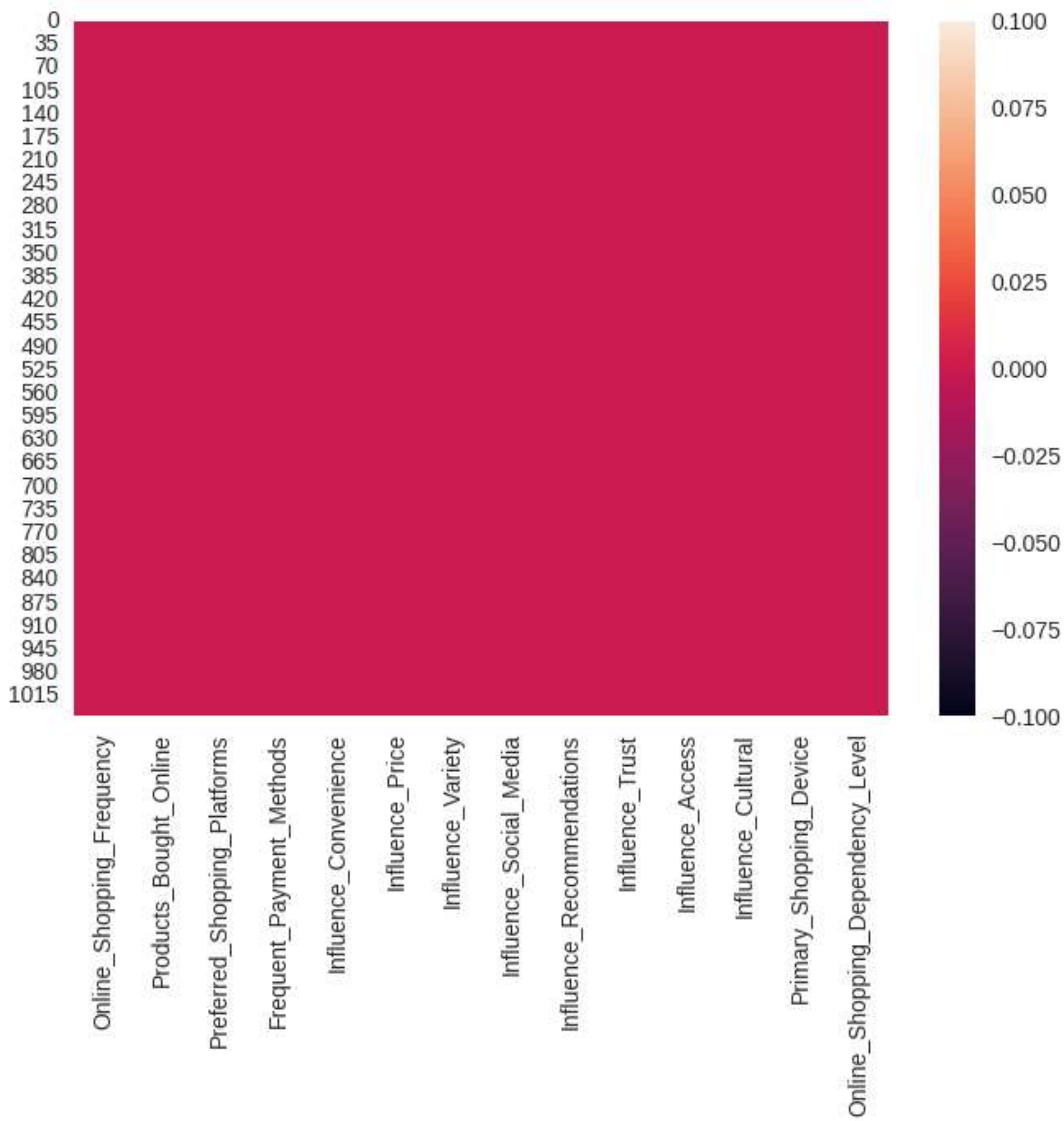

Fig. 6. Heatmap for Checking Null Value.

IV. Result and Discussion

This section includes an analysis of the results using performance evaluation metrics along with a tabular presentation of all experimental results for this investigation. Table I below displays the confusion matrix for each of the active classifiers. There has been a total of 262 instances of the test case in this experiment. Moreover, Table I shows the confusion matrix for the four working classifiers. From Table I it has been found that Logistic Regression, KNN, and SVC have the highest True Positive count which is 180 and MLP has the lowest True Positive Count which is 178. Again, the KNN classifier has the highest True Negative count among Logistic Regression, SVC, and MLP which is 73. On the other hand, the Logistic Regression classifier has the lowest True Negative count of the other three working classifiers which is 59.

Table II compares the performance of four working classifiers. From Table II, it has been observed that the classifier namely K-Nearest Neighbors (KNN) has outperformed than other three working classifiers and obtained the highest test accuracy of 0.97. Logistic Regression, Support Vector Classifier (SVC), and Multilayer Perceptron (MLP) classifiers have achieved test accuracy of 0.91, 0.95, and 0.93. If the data set does not have equal data of several types (for this experiment, such as High or Low which have been shown in Fig. 7.) and the cost of False Positive and False Negative are very different, then accuracy cannot justify the performance of the model properly [57]. In that case, it has been necessary to look into both the precision and recall which indicates to look into the F1 score [58]. The dataset that has been used in this work has more records of High compared to the Low which has been shown in Fig. 7. with the cost variation of False Positive and False

Negative. Particularly, from Fig. 7. it has been found that among 1045 respondents 69.86% of the people have considered their online shopping dependency level as high on the other hand, 30.14% of the people have considered their online shopping dependency level as low. In actuality, internet purchasing is increasing in developing nations [59]. This trend has emerged in recent years and has only become stronger [60]. Due to lockdowns, social isolation, and health concerns, the COVID-19 epidemic in particular has accelerated the expansion of online shopping in many developing countries [61]. As a result, people are increasingly turning to online markets for their purchasing needs [62]. The pandemic has highlighted the accessibility, simplicity, and security of online shopping [63].

The False Positive Rate of Logistic Regression, KNN, and SVC have been the same which is 0.00%. So, it has been important to look into Precision, Recall, F1-Score, and Specificity.

same Precision which is 0.95 and the MLP classifier has the highest False Positive Rate which is 1.11% but the Logistic Regression classifier has the lowest test accuracy among KNN, SVC, and MLP classifiers which is 0.91.

It is essential to compare training accuracy with testing accuracy in order to comprehend the overfitting and underfitting scenarios in a machine learning model [64]. It has been determined that this model is not overfitting or underfitting by comparing the train accuracy and test accuracy of each classifier in Table 2 analysis.

Fig. 8. has showed the comparison of testing accuracy of all four working algorithms in a bar graph form. A bar chart or bar graph uses rectangular bars with heights or lengths appropriate to the values they represent to show categorical data [65].

If a machine learning model consistently achieves high accuracy and performs well in terms of Precision, Recall, and

TABLE I. CONFUSION MATRIX OF ALL THE FOUR WORKING ALGORITHMS

Algorithm Name	True Positive	False Negative	False Positive	True Negative
Logistic Regression	180	0	23	59
K-Nearest Neighbors (KNN)	180	0	9	73
Support Vector Classifier (SVC)	180	0	13	69
Multilayer Perceptron (MLP)	178	2	16	66

TABLE II. COMPARISON OF FOUR WORKING ALGORITHMS BASED ON PERFORMANCE EVALUATION METRICS

Algorithm Name	Train Accuracy	Test Accuracy	Precision	Recall	F1-Score	Specificity	False Positive Rate (%)	False Negative Rate (%)
Logistic Regression	0.92	0.91	0.95	0.86	0.89	0.72	0.00	28.04
KNN	0.97	0.97	0.98	0.95	0.96	0.89	0.00	10.98
SVC	0.96	0.95	0.97	0.92	0.94	0.84	0.00	15.85
MLP	0.93	0.93	0.95	0.90	0.92	0.80	1.11	19.51

The Precision, Recall, F1-Score, and Specificity of the KNN classifier have been the highest than the other three classifiers which are 0.98, 0.95, 0.96, and 0.89 respectively. From Table II it has been also found that the KNN classifier has the lowest False Negative Rate which is 10.98%. On the other hand, Logistic Regression has the lowest Recall, F1-Score, and Specificity which are 0.86, 0.89, and 0.72. Also, Logistic Regression has the highest False Negative Rate which is 28.04%. Though the Logistic Regression and MLP have the

F1-Score, it is presumably appropriate for the task at hand [66]. After analyzing Fig. 8., it has been found that the best accuracy has been achieved by the K-Nearest Neighbors (KNN) which has been also proved by Table I and Table II. Finally, the K-Nearest Neighbor (KNN) has been chosen as the ultimate algorithm.

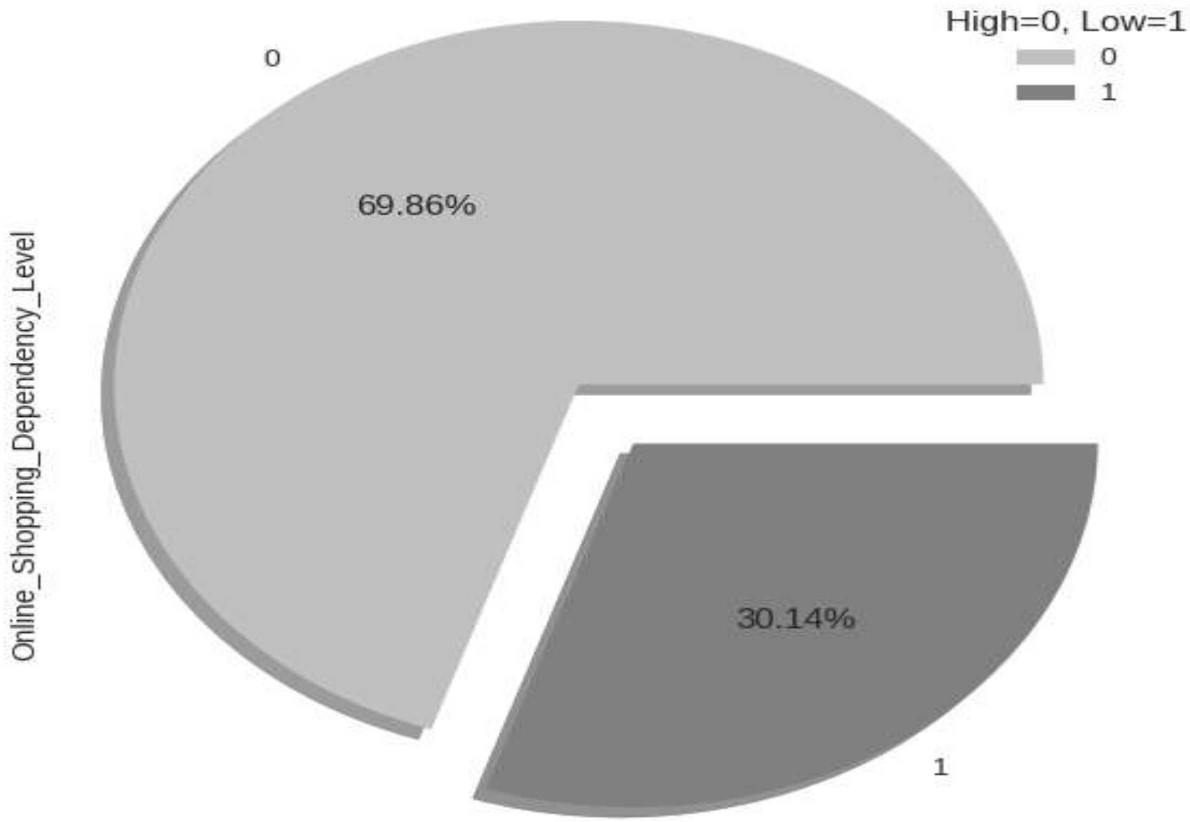

Fig. 7. People's Thoughts on Their Dependency in Online Shopping.

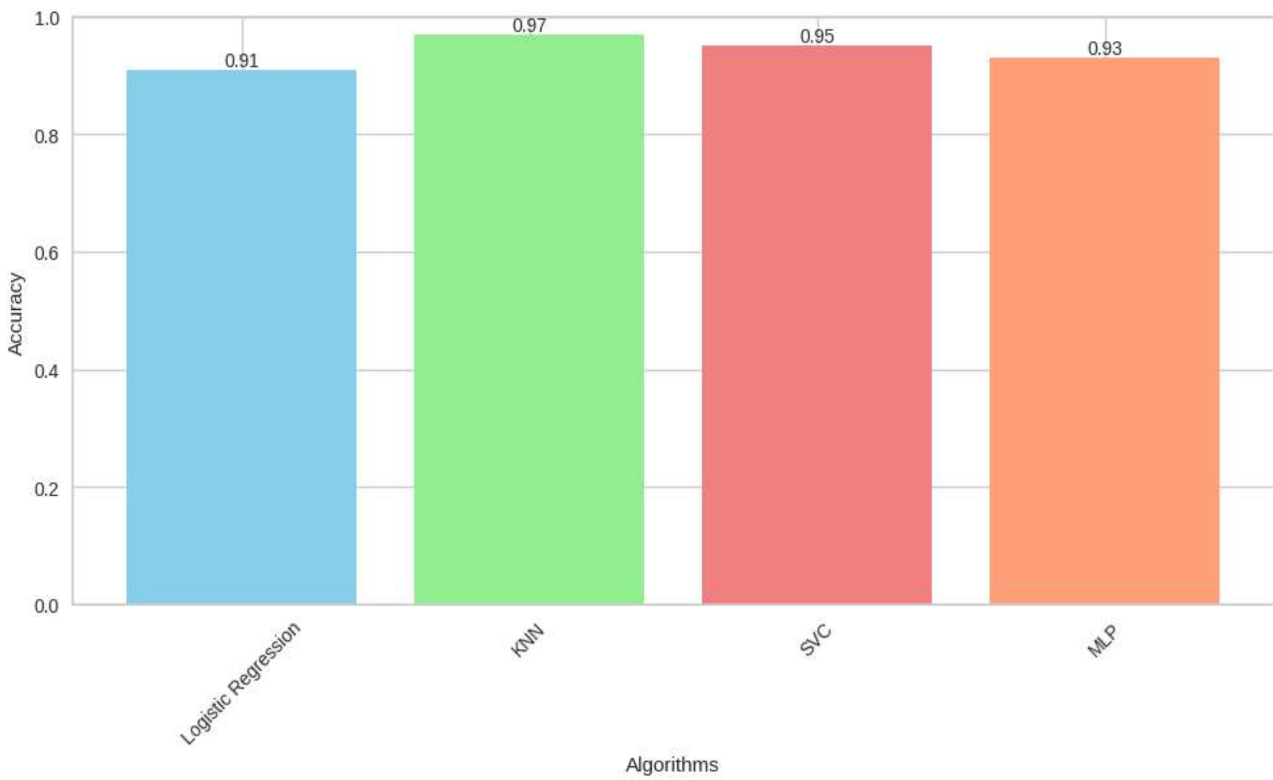

Fig. 8. Comparison of Testing Accuracy of Four Working Algorithms.

V. Conclusion

The outcome of this work has demonstrated the outstanding potential of machine learning techniques for predicting the levels of online shopping reliance among Bangladeshis. This investigation has led to several discoveries and insights that have illuminated the complex relationships between human behavior, culture, and technological growth. The dance that algorithms conducted on the data produced a tapestry of consumer preferences and behaviors, illuminating how traditional values and Bangladeshi culture's tremendous expansion in digital commerce coexist in harmony. It became clear how pervasive the effects of technology were on daily life, affecting people's routines and choices in both bustling cities and peaceful countryside regions. The K-Nearest Neighbors (KNN) classifier beats the other three in terms of predicting these dependency levels. Participants overwhelmingly self-reported having high degrees of dependency, which highlights how critical this issue is in Bangladesh's online shopping sector. This research explores uncharted territory in the disciplines of machine learning and dependency on online shopping, going beyond mere empirical observation. It is a spectacular illustration of the KNN's transformative ability and serves as a baseline for future studies and therapies. The dominance of KNN in predicting dependency levels highlights the need to know, manage, and resolve this issue in Bangladeshi society. This journey goes beyond analysis and model creation to examine the dynamic spirit of an emerging digital civilization. Anticipating dependency on online shopping may be used to foster empathy and change as well as increase understanding. It empowers customers, businesses, and officials alike by opening doors to the formulation of useful initiatives and the comprehension of existing issues.

VI. Future Work

In the future, more features will be added to the dataset with more people's data than the present dataset and a deep analysis will be conducted between the factors.

Acknowledgment

Appreciation is expressed to the respondents who helped to complete this research by sharing potential information.

References

[1] Helbing, D. Globally networked risks and how to respond. *Nature* 497, 51–59 (2013). https://doi.org/10.1038/nature12047

[2] Abusaleh, Kazi, M. Rezaul Islam, and Md Nurul Islam. *Globalization and urban culture in Dhaka*. Taylor & Francis, 2022. https://doi.org/10.4324/9781003096979

[3] Sudhir, K., Priester, J., Shum, M. *et al.* Research Opportunities in Emerging Markets: an Inter-disciplinary Perspective from Marketing, Economics, and Psychology. *Cust. Need. and Solut.* **2**, 264–276 (2015). https://doi.org/10.1007/s40547-015-0044-1

[4] Turow, Joseph. *The aisles have eyes: How retailers track your shopping, strip your privacy, and define your power*. Yale University Press, 2017. https://doi.org/10.12987/9780300225075

[5] Mpinganjira, Mercy. "Online store service convenience, customer satisfaction and behavioural intentions: a focus on utilitarian oriented shoppers." *Journal of Economics and Behavioral Studies* 7.1 (J) (2015): 36-49. https://doi.org/10.22610/jebs.v7i1(J).561

[6] Meuris, Jirs, and Carrie R. Leana. "The high cost of low wages: Economic scarcity effects in organizations." *Research in Organizational Behavior* 35 (2015): 143-158. https://doi.org/10.1016/j.riob.2015.07.001

[7] Darwish, S., Gomes, A. (2023). E-commerce and Impact of COVID-19 on Consumer Behaviors Globally: A Review. In: Musleh Al-Sartawi, A.M.A., Razzaque, A., Kamal, M.M. (eds) From the Internet of Things to the Internet of Ideas: The Role of Artificial Intelligence. EAMMIS 2022. Lecture Notes in Networks and Systems, vol 557. Springer, Cham. https://doi.org/10.1007/978-3-031-17746-0_36

[8] H. Wang, C. Ma and L. Zhou, "A Brief Review of Machine Learning and Its Application," *2009 International Conference on Information Engineering and Computer Science*, Wuhan, China, 2009, pp. 1-4, doi: 10.1109/ICIECS.2009.5362936.

[9] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

[10] Trivedi, S.K., Patra, P., Srivastava, P.R. *et al.* What prompts consumers to purchase online? A machine learning approach. *Electron Commer Res* (2022). https://doi.org/10.1007/s10660-022-09624-x

[11] Aziz, Abdul, and Umma Naima. "Rethinking digital financial inclusion: Evidence from Bangladesh." *Technology in Society* 64 (2021): 101509. https://doi.org/10.1016/j.techsoc.2020.101509

[12] Bieron, Brian, and Usman Ahmed. "Regulating e-commerce through international policy: Understanding the international trade law issues of e-commerce." *Journal of World Trade* 46.3 (2012). https://doi.org/10.54648/trad2012017

[13] Cai, Y., Cude, B.J. (2016). Online Shopping. In: Xiao, J. (eds) Handbook of Consumer Finance Research. Springer, Cham. https://doi.org/10.1007/978-3-319-28887-1_28

[14] Lohse, G. L., Bellman, S., & Johnson, E. J. (2000). Consumer Buying Behavior on the Internet: Findings from Panel Data. *Journal of Interactive Marketing*, *14*(1), 15-29. https://doi.org/10.1002/(SICI)1520-6653(200024)14:1<15::AID-DIR2>3.0.CO;2-C

[15] Shaffer, Greg, and Z. John Zhang. "Competitive coupon targeting." *Marketing Science* 14.4 (1995): 395-416. https://doi.org/10.1287/mksc.14.4.395

[16] Duffy, Dennis L. "Affiliate marketing and its impact on e-commerce." *Journal of consumer marketing* 22.3 (2005): 161-163. doi:10.1108/07363760510595986

[17] Rust, Roland T., and Tuck Siong Chung. "Marketing models of service and relationships." *Marketing science* 25.6 (2006): 560-580. https://doi.org/10.1287/mksc.1050.0139

[18] Libai, Barak, Eyal Biyalogorsky, and Eitan Gerstner. "Setting referral fees in affiliate marketing." *Journal of Service Research* 5.4 (2003): 303-315. https://doi.org/10.1177/1094670503005004003

[19] Homburg, Christian, Mathias Droll, and Dirk Totzek. "Customer prioritization: does it pay off, and how should it be implemented?." *Journal of Marketing* 72.5 (2008): 110-130. https://doi.org/10.1509/jmkg.72.5.110

[20] Chen, Y., Ghosh, A., McAfee, R.P., Pennock, D. (2008). Sharing Online Advertising Revenue with Consumers. In: Papadimitriou, C., Zhang, S. (eds) Internet and Network Economics. WINE 2008. Lecture Notes in Computer Science, vol 5385. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-92185-1_62

[21] Jahng, Jungjoo, Hemant Jain, and Keshavamurthy Ramamurthy. "Effects of interaction richness on consumer attitudes and behavioral intentions in e-commerce: Some experimental results." *European Journal of Information Systems* 16.3 (2007): 254-269. https://doi.org/10.1057/palgrave.ejis.3000665

[22] Manns, Jeffrey. "The Case for Preemptive Oligopoly Regulation." *Ind. LJ* 96 (2020): 751.

[23] Berg, Natalie, and Miya Knights. *Amazon: How the world's most relentless retailer will continue to revolutionize commerce*. Kogan Page Publishers, 2021.

[24] Altinkemer, K., Ozcelik, Y. Cash-back rewards versus equity-based electronic loyalty programs in e-commerce. *Inf Syst E-Bus Manage* **7**, 39–55 (2009). https://doi.org/10.1007/s10257-007-0062-0

[25] Fu, Tao, et al. "Percolation on shopping and cashback electronic commerce networks." *Physica A: Statistical Mechanics and its Applications* 392.12 (2013): 2807-2820. https://doi.org/10.1016/j.physa.2013.02.018

[26] Ho, Yi-Chun Chad, Yi-Jen Ian Ho, and Yong Tan. "Online cashback pricing: a new affiliate strategy for e-business." *Proceedings of the 34th international conference on information systems (ICIS)*. 2013.

[27] Roig-Tierno, N., Kraus, S. & Cruz, S. The relation between coopetition and innovation/entrepreneurship. *Rev Manag Sci* 12, 379–383 (2018). https://doi.org/10.1007/s11846-017-0266-8

[28] Vana, Prasad, Anja Lambrecht, and Marco Bertini. "Cashback is cash forward: delaying a discount to entice future spending." *Journal of Marketing Research* 55.6 (2018): 852-868. https://doi.org/10.1177/0022243718811853

[29] Ballestar, María Teresa, Pilar Grau-Carles, and Jorge Sainz. "Consumer behavior on cashback websites: Network strategies." *Journal of Business Research* 69.6 (2016): 2101-2107. https://doi.org/10.1016/j.jbusres.2015.12.015

[30] Ballestar, María Teresa, Jorge Sainz, and Joan Torrent-Sellens. "Social networks on cashback websites." *Psychology & Marketing* 33.12 (2016): 1039-1045. https://doi.org/10.1002/mar.20937

[31] Ballestar, María Teresa, Pilar Grau-Carles, and Jorge Sainz. "Customer segmentation in e-commerce: Applications to the cashback business model." *Journal of Business Research* 88 (2018): 407-414. https://doi.org/10.1016/j.jbusres.2017.11.047

[32] Monaghan, Sinéad, Patrick Gunnigle, and Jonathan Lavelle. "Firm-location dynamics and subnational institutions: creating a framework for collocation advantages." *Industry and Innovation* 25.3 (2018): 242-263. https://doi.org/10.1080/13662716.2017.1315562

[33] Lamberton, Cait, and Andrew T. Stephen. "A thematic exploration of digital, social media, and mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry." *Journal of marketing* 80.6 (2016): 146-172. https://doi.org/10.1509/jm.15.0415

[34] Eggers, Fabian, et al. "Technologies that support marketing and market development in SMEs—Evidence from social networks." *Journal of small business management* 55.2 (2017): 270-302. https://doi.org/10.1111/jsbm.12313

[35] Kumar, Vijay, and Shaphali Gupta. "Conceptualizing the evolution and future of advertising." *Journal of advertising* 45.3 (2016): 302-317. https://doi.org/10.1080/00913367.2016.1199335

[36] Harmeling, C.M., Moffett, J.W., Arnold, M.J. *et al.* Toward a theory of customer engagement marketing. *J. of the Acad. Mark. Sci.* 45, 312–335 (2017). https://doi.org/10.1007/s11747-016-0509-2

[37] Md. Safiul Alam, Anirban Roy, Partha Protim Majumder and Sharun Akter Khushbu, “Predicting the Level of Safety Feeling of Bangladeshi Internet users using Data Mining and Machine Learning” International Journal of Advanced Computer Science and Applications(IJACSA), 14(9), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140976

[38] Vafeiadis, Thanasis, et al. "A comparison of machine learning techniques for customer churn prediction." *Simulation Modelling Practice and Theory* 55 (2015): 1-9. https://doi.org/10.1016/j.simpat.2015.03.003

[39] Kramer, O. (2013). K-Nearest Neighbors. In: Dimensionality Reduction with Unsupervised Nearest Neighbors. Intelligent Systems Reference Library, vol 51. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38652-7_2

[40] Cunningham, Padraig, and Sarah Jane Delany. "k-Nearest neighbour classifiers-A Tutorial." *ACM computing surveys (CSUR)* 54.6 (2021): 1-25. https://doi.org/10.1145/3459665

[41] Bansal, Malti, Apoorva Goyal, and Apoorva Choudhary. "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning." *Decision Analytics Journal* 3 (2022): 100071. https://doi.org/10.1016/j.dajour.2022.100071

[42] Ghiasi, Alireza, Ching-Tai Ng, and Abdul Hamid Sheikh. "Damage detection of in-service steel railway bridges using a fine k-nearest neighbor machine learning classifier." *Structures*. Vol. 45. Elsevier, 2022. https://doi.org/10.1016/j.istruc.2022.10.019

[43] Lau, K. W., and Q. H. Wu. "Online training of support vector classifier." *Pattern Recognition* 36.8 (2003): 1913-1920. https://doi.org/10.1016/S0031-3203(03)00038-4

[44] Tehrany, M.S., Jones, S., Shabani, F. *et al.* A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theor Appl Climatol* 137, 637–653 (2019). https://doi.org/10.1007/s00704-018-2628-9

[45] S. Ghosh, A. Dasgupta and A. Swetapadma, "A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification," *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, Palladam, India, 2019, pp. 24-28, doi: 10.1109/ISS1.2019.8908018.

[46] Westreich, Daniel, Justin Lessler, and Michele Jonsson Funk. "Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression." *Journal of clinical epidemiology* 63.8 (2010): 826-833. https://doi.org/10.1016/j.jclinepi.2009.11.020

[47] Tharwat, A. Parameter investigation of support vector machine classifier with kernel functions. *Knowl Inf Syst* 61, 1269–1302 (2019). https://doi.org/10.1007/s10115-019-01335-4

[48] Smits, J. R. M., et al. "Using artificial neural networks for solving chemical problems: Part I. Multi-layer feed-forward networks." *Chemometrics and Intelligent Laboratory Systems* 22.2 (1994): 165-189. https://doi.org/10.1016/0169-7439(93)E0035-3

[49] J. Tang, C. Deng and G. -B. Huang, "Extreme Learning Machine for Multilayer Perceptron," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 4, pp. 809-821, April 2016, doi: 10.1109/TNNLS.2015.2424995.

[50] Malalur, Sanjeev S., Michael T. Manry, and Praveen Jesudhas. "Multiple optimal learning factors for the multi-layer perceptron." *Neurocomputing* 149 (2015): 1490-1501. https://doi.org/10.1016/j.neucom.2014.08.043

[51] Poulton, Mary M. "Multi-layer perceptrons and back-propagation learning." *Handbook of Geophysical Exploration: Seismic Exploration*. Vol. 30. Pergamon, 2001. 27-53. https://doi.org/10.1016/S0950-1401(01)80017-3

[52] Chicco, Davide, Niklas Tötsch, and Giuseppe Jurman. "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation." *BioData mining* 14.1 (2021): 1-22.doi: 10.1186/s13040-021-00244-z

[53] Sokolova, Marina, and Guy Lapalme. "A systematic analysis of performance measures for classification tasks." *Information processing & management* 45.4 (2009): 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

[54] DeVries, Zachary, et al. "Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability." *The spine journal* 21.7 (2021): 1135-1142. https://doi.org/10.1016/j.spinee.2021.02.007

[55] Chicco, Davide, and Giuseppe Jurman. "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation." *BMC genomics* 21.1 (2020): 1-13.doi: 10.1186/s12864-019-6413-7

[56] Matos, Luís Miguel, et al. "Categorical attribute transformation environment (cane): A python module for categorical to numeric data preprocessing." *Software Impacts* 13 (2022): 100359. https://doi.org/10.1016/j.simpa.2022.100359

[57] M. V. Joshi, V. Kumar and R. C. Agarwal, "Evaluating boosting algorithms to classify rare classes: comparison and improvements," *Proceedings 2001 IEEE International Conference on Data Mining*, San Jose, CA, USA, 2001, pp. 257-264, doi: 10.1109/ICDM.2001.989527.

[58] Provost, Foster J., Tom Fawcett, and Ron Kohavi. "The case against accuracy estimation for comparing induction algorithms." *ICML*. Vol. 98. 1998.

[59] Kamssu, Aurore J., et al. "Shortcomings to globalization: Using Internet technology and electronic commerce in developing countries." *The Journal of Developing Areas* (2004): 151-169.

[60] Alyoubi, Adel A. "E-commerce in developing countries and how to develop them during the introduction of modern systems." *Procedia Computer Science* 65 (2015): 479-483. https://doi.org/10.1016/j.procs.2015.09.127

[61] Almajali, Dmaithan Abdelkarim, and Qais Hammouri. "Predictors of online shopping during Covid-19 pandemic in developing country: Qualitative analysis." *Annals of the Romanian Society for Cell Biology* 25.6 (2021): 12970-12977.

[62] Pookulangara, Sanjukta, and Kristian Koesler. "Cultural influence on consumers' usage of social networks and its' impact on online purchase intentions." *Journal of Retailing and Consumer Services* 18.4 (2011): 348-354. https://doi.org/10.1016/j.jretconser.2011.03.003

[63] Paștiu, C.A.; Oncioiu, I.; Gârdan, D.A.; Maican, S.Ș.; Gârdan, I.P.; Muntean, A.C. The Perspective of E-Business Sustainability and Website Accessibility of Online Stores. *Sustainability* 2020, *12*, 9780. https://doi.org/10.3390/su12229780

[64] Montesinos López, O.A., Montesinos López, A., Crossa, J. (2022). Overfitting, Model Tuning, and Evaluation of Prediction Performance. In: Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer, Cham. https://doi.org/10.1007/978-3-030-89010-0_4

[65] Cooper, Linda L., and Felice S. Shore. "The effects of data and graph type on concepts and visualizations of variability." *Journal of statistics education* 18.2 (2010). https://doi.org/10.1080/10691898.2010.11889487

[66] Van Atteveldt, Wouter, Mariken ACG Van der Velden, and Mark Boukes. "The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms." *Communication Methods and Measures* 15.2 (2021): 121-140. https://doi.org/10.1080/19312458.2020.1869198